CLUSTER ANALYSIS OF THE EFFECTIVENESS OF MANAGEMENT OF HIGHER EDUCATION INSTITUTIONS

Aidar Mambetkaziyev  
Rector*  
Zhasultan Baikenov  
Corresponding author  
First Vice-President*  
E-mail: zhass.baiken@yahoo.com  
Galina Konopyanova  
PhD, Professor*  
*Department of Business  
Kazakh-American Free University  
M. Gorkiy str., 76, Ust-Kamenogorsk, Republic of Kazakhstan, 070000

Keywords: cluster analysis, k-means method, agglomerative cluster analysis, university management level

1. Introduction

Each ranking system is based on a certain group of indicators reflecting certain aspects of the quality of the university operation [1–3]. These indicators are the result of the organization of the relevant activities of the higher educational institution [4]. As known, management has a direct and significant impact on the university's position in the ranking of universities. In other words, a high position in ranking tables is the result of management in the corresponding category of the university's activity [5–7].

The increase in competitiveness of a higher education institution is possible with an effective management organization in all areas of its activities. As noted above, the quality of university management is directly reflected in the ranking indicators [8–10]. Since 2008, the Independent Agency for Quality Assurance in Education (IQAA Ranking) publishes the results of the ranking of the best higher education institutions, which is based on international standards for ranking higher educational institutions [11]. The multidimensional National University Ranking was scientifically substantiated and approved by international experts in the field of higher education at the conferences of the International Ranking Expert Group (IREG) in Shanghai and Bratislava, was published in the scientific journals “Higher Education in Europe” (London) [1] and “Journal of Higher Education” (Shanghai) [2], reviewed in the book “Rankings and the Reshaping of Higher Education: the Battle for World Wide Excellence” [3].

In the articles [12–14], researchers proposed to assess university management based on key indicators that most objectively reflect the quality of its operation: quality of research and development, quality of teaching staff, quality of innovations, commercialization of research results, quality of facilities, research laboratories, quality of teaching methods and research technologies, quality of academic mobility and cooperation, etc. To cluster universities, they used a method of data simulation. In the papers [15–17], the k-means method to cluster students from four universities by their academic performance and behavior was used. In turn, the work [18] illustrated the use of k-means clustering to analyze the characteristics of learning behavior when students are engaged in problem solving in an online learning environment.

Based on the ranking indicators of certain universities, it is possible to assess the level of organization of management of the corresponding activity. Therefore, the research of clustering the effectiveness of management of higher education institutions is relevant.

2. Literature review and problem statement

The paper [19] testifies that the issues of assessing the effectiveness of management in organizations have not only important theoretical, but also practical significance. A system consisting of three categories to assess the performance of some universities, based on indicators of the effectiveness...
of three main categories was used: the first category — indicators of graduates and attracting funds for research; the second category is the qualification of graduates and their readiness to work, employee publications, patents; and the third category is expenditures, student-faculty ratio, and faculty workload. However, due to the small sample size, this study did not conduct a comparative analysis of management performance evaluation in universities in different countries. In turn, the article [20] assessed 19 ranking systems in Australia, Spain, China, Canada, the USA, and other countries. The results of this research show that despite the differences in geographic location and culture, rankings reveal the best educational institutions, although the assessment needs to be supplemented with other indicators.

In [21], researchers proposed a model for measuring the performance of university research management, based on a balanced assessment of quantitative indicators such as finance, customers, innovation and learning, internal business, alliances, and networks. In turn, in [22], a university ranking based on a hybrid multi-criterion decision-making model (MCDM) was conducted. Evaluating the ranking results of 12 private universities, they tried to identify ways to improve university efficiency. But this problem was not completely solved because of the same type of all universities. In [23], an integrated approach to assessing the effectiveness of management in universities, based on an integral index covering individual management parameters was used. Despite the fact that these parameters of university management were evaluated as an integrated result of individual, group and organizational activities based on synergistic effect, they were not summarized in the index.

The work [24] proposed to evaluate the effectiveness of the management system by evaluating a number of indicators, such as administration efficiency, administrative staff turnover, administration development, personnel competency, the coefficient of strategic reliability, the level of criticism of managerial actions, the number of established management methods (issued orders, guidelines, instructions, regulations, tariffs, budgets, etc.), management leadership index. But such an indicator as the coefficient of settlement and prevention of dysfunctional conflicts was not taken into account.

In the paper [25], the researchers identified three groups of indicators, according to which, in their opinion, the effectiveness of university management should be measured: assessment of university administrative staff; assessment of the performance of certain management departments; assessment of the university management system. In addition, the effectiveness of higher education institutions on the basis of a multilevel fuzzy model, which was divided into three levels of management, was assessed. Each level of management had a corresponding group of factors reflecting the degree of management: group of factors I — operational level; group of factors II — tactical management level; group of factors III — strategic management. However, in this research, the problem of a balanced scorecard remained unresolved.

In turn, the methodology presented in [26] describes the use of a balanced scorecard that allows a comprehensive assessment of the effectiveness and efficiency of the university management system. It should be noted that the problem of determining the financial indicators of the university’s activities and assessing its effectiveness has not been solved, because this is the greatest difficulty when using a balanced scorecard.

Literature analysis shows that there are many works that use various methods to assess university management, most of which are presented in the form of integral indicators. However, there are not so many works on the assessment of individual areas of activity in universities using clustering methods. Moreover, clustering universities according to the criteria of the national university ranking system in order to assess university management has never been done before.

3. The aim and objectives of the study

The aim of this study is to cluster the efficiency of management of higher education institutions on the example of the Republic of Kazakhstan. This will make it possible to determine the stages and levels of development of management in universities.

To achieve the aim, the following objectives were set:
- to explore the multidimensional ranking and features of the clear clustering method;
- to determine the distribution of universities according to the fuzzy clustering method;
- to define the centroids of clusters based on the agglomerative cluster analysis method.

4. Materials and methods

The object of the study is the internal structure of management in universities. The hypothesis of the research: the indicators (quality of education, indicators of employment of university graduates, the demand for the graduates in the labor market, the symbiosis of science, education and business, and mobility of students) are a direct result of effective management in universities. The assumption of the study is that the clustering of universities makes it possible to determine consistency in relation to the organization of university management. The simplification of the study is that the division of leading universities into clusters does not contribute to the analysis of the internal structure of management.

To conduct cluster analysis, three clustering algorithms will be used to ensure the accuracy of results: the clear k-means method, the fuzzy k-means method, and the agglomerative hierarchical clustering (AHC) method. The k-means clustering algorithm is the traditional clustering algorithm proposed by McQueen, which is simple and efficient [27]. At the same time, it has the advantages of scalability and high efficiency for processing large datasets [28]. The k-means clustering algorithm has a wide range of applications [29].

This method breaks a set of elements of the vector space into a predetermined number of clusters k. The essence of the algorithm is that it seeks to minimize the standard deviation at the points of each cluster [30]. The main idea of this method is that, at each iteration, the centroid for each cluster obtained at the previous step is recalculated, then the vectors are divided into clusters again in accordance with which of the new centers is closer according to the chosen metric [31, 32]. The algorithm ends when no cluster changes occur at a certain iteration.

Agglomerative cluster analysis (AHC) is a bottom-up approach in which each observation starts in its own cluster and pairs of clusters are combined with the advancement up the hierarchy. In our analysis, Euclidean distance is taken as a metric, and Ward's criterion is taken as a criterion to determine the relationship between observation sets A and B. In our case, having a certain ranking according to
the relevant indicators, we are supposed to divide universities according to the selected characteristics into 5 main clusters and highlight the most effective management tools for each cluster. It should be noted that the differences between clusters should be obvious, and within a cluster, university indicators should be as similar as possible.

5. Results of cluster formation

5.1. Features of the clear clustering method
Currently, the multidimensional ranking includes 7 academic indicators (indicator 1 – diversity of the student population; indicator 2 – student learning outcomes; indicator 3 – academic staff; indicator 4 – research and development and innovative work; indicator 5 – international cooperation; indicator 6 – informational provision; indicator 7 – graduates’ employment) and 3 reputation assessments: by experts and employers, by current university students and by university graduates.

The academic performance of universities is 80 % and the reputation score is 20 % of the total score. The distribution of these 80 % in the ranking is as follows (Fig. 1).

Using the clear clustering method, the following results were obtained (Table 1).

In Table 1, the data show the centers of the clusters around which the rest of the universities belonging to a particular cluster are grouped according to the ranking by the corresponding indicators. Based on the data obtained [33], it can be noted that the center of cluster 1 is Al-Farabi KazNU, cluster 2 – Kazakh National Women’s Pedagogical University, cluster 3 – Kazakh American Free University, cluster 4 – S. Toraigyrov Pavlodar State University and cluster 5 – Caspian Public University.

All the results obtained are statistically significant, except for clusters 3 and 5, for which the p-value is greater than 0.0001. Errors are 11.8 % and 16.2 %, respectively. The presence of such errors is associated with the university website ranking, according to which there were universities that were rather low in all previous ranking tables, but high in website ranking, and vice versa. At the same time, these errors do not diminish the significance of the model.

The results of cluster formation, as well as K-means method clustering indicators, are presented in Table 2. The results show that cluster 1 includes 2 universities, cluster 2 – 13 universities, cluster 3 – 5 universities, cluster 4 – 14 universities, and cluster 5 – 9 universities. At the same time, the smallest differences in the parameters of the cluster are characteristic of cluster 5 (4.173), the largest – cluster 1 (10.247).

Table 2 clearly demonstrates a high degree of dispersion across cluster 1 and across cluster 2. For example, in terms of academic indicators, a university may be in the top position and have the highest values, and in terms of employer ranking, it may be almost at the bottom of the list.

5.2. Peculiarities of the distribution of universities according to the fuzzy clustering method
Let us consider the results of using the fuzzy clustering method. Calculation indicators are presented in Table 3.

As can be seen from the data obtained in Table 3, the largest number of universities belongs to the second cluster and the smallest number to cluster 3. At the same time, the largest discrepancies between the parameters within the clusters are characteristic for cluster 4 (6.201), followed by cluster 2 (5.627). When comparing the discrepancies within

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Size</th>
<th>Minimum distance to centroid</th>
<th>Maximum distance to centroid</th>
<th>Average distance to centroid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>6</td>
<td>253.548</td>
<td>2.890</td>
<td>12.563</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>9</td>
<td>261.676</td>
<td>2.418</td>
<td>6.164</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>11</td>
<td>506.620</td>
<td>2.538</td>
<td>13.174</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>16</td>
<td>521.496</td>
<td>1.205</td>
<td>11.154</td>
</tr>
</tbody>
</table>

As can be seen from the data obtained in Table 3, the largest number of universities belongs to the second cluster and the smallest number to cluster 3. At the same time, the largest discrepancies between the parameters within the clusters are characteristic for cluster 4 (6.201), followed by cluster 2 (5.627). When comparing the discrepancies within
the clusters with the previous method, it can be noted that with such a distribution, these discrepancies are approximately the same. The distribution of universities by cluster is presented in Table 4.

<table>
<thead>
<tr>
<th>Indicators, %</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
<th>Cluster 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic resources</td>
<td>80.000</td>
<td>79.308</td>
<td>66.523</td>
<td>52.619</td>
<td>39.049</td>
</tr>
<tr>
<td>Expert assessment</td>
<td>5.000</td>
<td>4.882</td>
<td>4.307</td>
<td>3.991</td>
<td>3.343</td>
</tr>
<tr>
<td>Students</td>
<td>3.970</td>
<td>4.767</td>
<td>4.57</td>
<td>4.054</td>
<td>3.765</td>
</tr>
<tr>
<td>Graduates</td>
<td>4.830</td>
<td>4.378</td>
<td>4.151</td>
<td>3.877</td>
<td>3.420</td>
</tr>
<tr>
<td>Website rank</td>
<td>52.530</td>
<td>19.607</td>
<td>17.039</td>
<td>13.511</td>
<td>12.098</td>
</tr>
</tbody>
</table>

According to this distribution, 1 university belongs to cluster 1, 6 universities of the Republic of Kazakhstan are included in cluster 2, 9 universities belong to cluster 3, 11 universities are in cluster 4 and 16 universities are in cluster 5. In general, a clear grouping of universities can be noted, with the exception of a fairly strong scatter of values in clusters 2 and 4.

5. 3. Centroids of clusters based on the agglomerative cluster analysis method

Let us now consider the option of clustering using the agglomerative cluster analysis method. The summary statistics for this clustering method are presented in Table 5.

Table 5

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observations</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic resources</td>
<td>43</td>
<td>30.750</td>
<td>80.000</td>
<td>54.841</td>
<td>15.438</td>
</tr>
<tr>
<td>Expert assessment</td>
<td>43</td>
<td>1.660</td>
<td>5.000</td>
<td>3.963</td>
<td>0.812</td>
</tr>
<tr>
<td>Employers</td>
<td>43</td>
<td>2.690</td>
<td>5.000</td>
<td>3.995</td>
<td>0.756</td>
</tr>
<tr>
<td>Students</td>
<td>43</td>
<td>2.900</td>
<td>5.000</td>
<td>4.152</td>
<td>0.827</td>
</tr>
<tr>
<td>Graduates</td>
<td>43</td>
<td>1.980</td>
<td>5.000</td>
<td>3.877</td>
<td>0.827</td>
</tr>
<tr>
<td>Website rank</td>
<td>43</td>
<td>7.460</td>
<td>52.530</td>
<td>15.482</td>
<td>8.007</td>
</tr>
</tbody>
</table>

Table 6

<table>
<thead>
<tr>
<th>Class</th>
<th>Academic resources</th>
<th>Expert assessment</th>
<th>Employers</th>
<th>Students</th>
<th>Graduates</th>
<th>Website ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80.000</td>
<td>5.000</td>
<td>4.910</td>
<td>3.970</td>
<td>4.830</td>
<td>52.530</td>
</tr>
<tr>
<td>2</td>
<td>77.427</td>
<td>4.824</td>
<td>4.774</td>
<td>4.591</td>
<td>4.297</td>
<td>20.830</td>
</tr>
<tr>
<td>3</td>
<td>66.571</td>
<td>4.280</td>
<td>4.248</td>
<td>4.700</td>
<td>4.194</td>
<td>15.648</td>
</tr>
<tr>
<td>5</td>
<td>38.346</td>
<td>3.361</td>
<td>3.669</td>
<td>3.783</td>
<td>3.460</td>
<td>11.529</td>
</tr>
</tbody>
</table>

Within-class variance 0.000 77.790 17.215 62.532 25.118
Minimum distance to centroid 0.000 3.801 1.737 3.222 1.067
Average distance to centroid 0.000 7.454 3.682 7.064 4.283
Maximum distance to centroid 0.000 13.518 5.820 12.951 8.781

This diagram shows that cluster indicators are grouped to a greater extent by indicators of expert assessments of graduates, employers, students, and to a lesser extent by academic resources. Fig. 2 illustrates the grouping of universities according to the respective clusters. It also shows the discrepancy in the number of universities included in a given cluster.
6. Discussion of the results of cluster formation

The analysis made it possible to develop a mechanism for dividing universities using three clustering models into groups based on the ranking parameters (on the example of the Republic of Kazakhstan universities). We have identified 5 clusters according to the degree of decline in the level of management in these universities (Tables 1, 3).

The first cluster includes universities with an excellent level of management, which is reflected in the indicators of academic resources and the indicators of graduates’ employment. At the same time, these universities have a high reputation among employers and students. Working for the external environment, university management also pays much attention to the work of the universities’ websites. The second cluster includes universities with a high level of management, which is reflected in the indicators of academic resources. However, these institutions may not have a high reputation among employers (Table 2).

The third cluster includes universities with an average level of management. These universities, as a rule, have average indicators of academic resources, average level of the university reputation among employers, average indicators of employment. The fourth cluster includes universities with a low level of management. These universities are characterized by average indicators of academic resources. But the problem is that the management of the university is not able to organize these resources and direct them to increase the level of student employment, to increase the university’s reputation among employers.

Only 1 university of the Republic of Kazakhstan could not be identified as belonging to any cluster, since its indicators correspond to the possibility of distribution among several clusters according to the three methods of analysis used (Fig. 2). This is M. Kozybayev North Kazakhstan University, which can be assigned to cluster 3 based on the clear clustering method, to cluster 4 based on the fuzzy clustering method, and to cluster 5 based on the AHC method. The lack of consistency in the analysis results made it impossible to assign the university to a certain cluster.

The limitations of the study lie in the fact that only three clustering models were developed (clear and fuzzy clustering based on k-means and agglomerative cluster analysis). The disadvantages of the study are that the study of the clustering of the management efficiency of higher educational institutions was conducted only in the Republic of Kazakhstan. Prospects for future research would include a larger sample of countries and universities. Further work should also focus on the study of the correlation between the indicated university ranking indicators and the level of university management.

7. Conclusions

1. The results of the clear clustering method demonstrates a high degree of dispersion across cluster 1 and cluster 2. This is largely due to the difficulty of finding and establishing uniform or roughly similar parameters for including a university in a particular cluster, since, as the ranking data show, universities have versatile indicators in different rankings.

2. When using the fuzzy clustering method, the difficulty arises due to the fact that since universities can have approximately the same values, they can be assigned to several clusters simultaneously. Therefore, the fuzzy clustering method can distribute universities according to the most appropriate clustering parameters.

3. In general, considering the centroids of the clusters, it can be noted that when performing a distribution based on the AHC method, a certain gradation of indicators is observed. It was stated there is a serious discrepancy in the sizes of the clusters formed.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Financing

The study was performed without financial support.

Data availability

The data that support the findings of this study are available on request from the corresponding author.

References

2. The Journal of Higher Education. Available at: https://www.tandfonline.com/toc/uhej20/64/5?nav=tocList
33. Independent Agency for Quality Assurance in Education. Available at: https://iqa-ranking.kz/en/